Improving 2D Diffusion Models for 3D Medical Imaging with Inter-Slice Consistent Stochasticity

Chenhe Du[1], Qing Wu[1], Xuanyu Tian[1], Jingyi Yu[1], Hongjiang Wei[2], Yuyao Zhang[1]

[1] ShanghaiTech University [2] Shanghai Jiao Tong University

Smile Lab — Smart Medical Image Learning And Enhancement Lab

ICLR

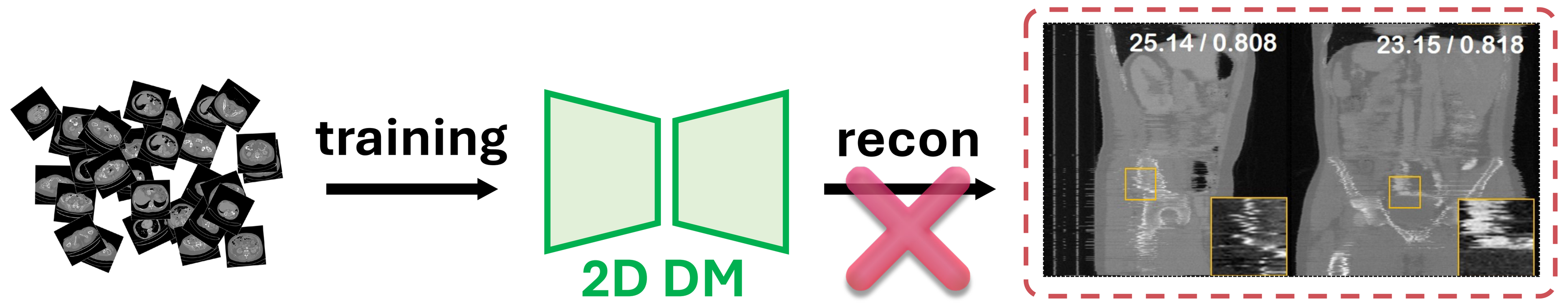
CONTACT ME

Background

- **Curse of dimensionality**: Training diffusion models (DMs) directly on 3D volumetric data is challenging.

× High dimensionality
× heavy computational cost
× insufficient data.

- **Realistic and practical solution**: Train on 2D slices, denoise slice by slice, then stack into a volume.



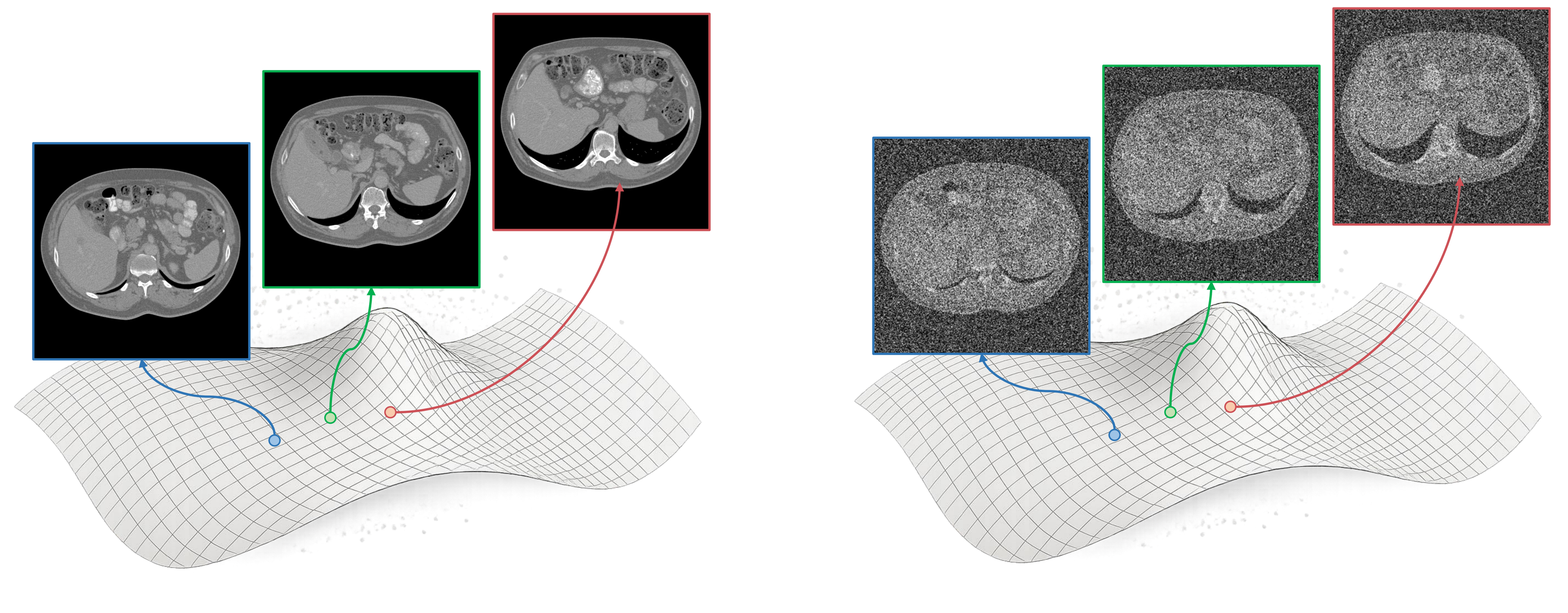
inter-slice discontinuities

- **Intrinsic randomness**: The stochasticity of diffusion sampling causes independent slice-wise sampling to produce inter-slice discontinuities.
- **Solution**: ***I**nter-**S**lice **C**onsistent **S**tochasticity (ISCS)*: Control inter-slice stochasticity to align sampling trajectories and mitigate inter-slice discontinuities *at the source*.

Motivation

Intuition: The core idea is straightforward.

- **If two slices are adjacent in the clean data manifold, they should also be close in the noise manifold.**
- Therefore, **their stochastic noise during diffusion sampling should be strongly correlated, rather than fully independent.**



Clean data manifold **Noisy data manifold**

Method

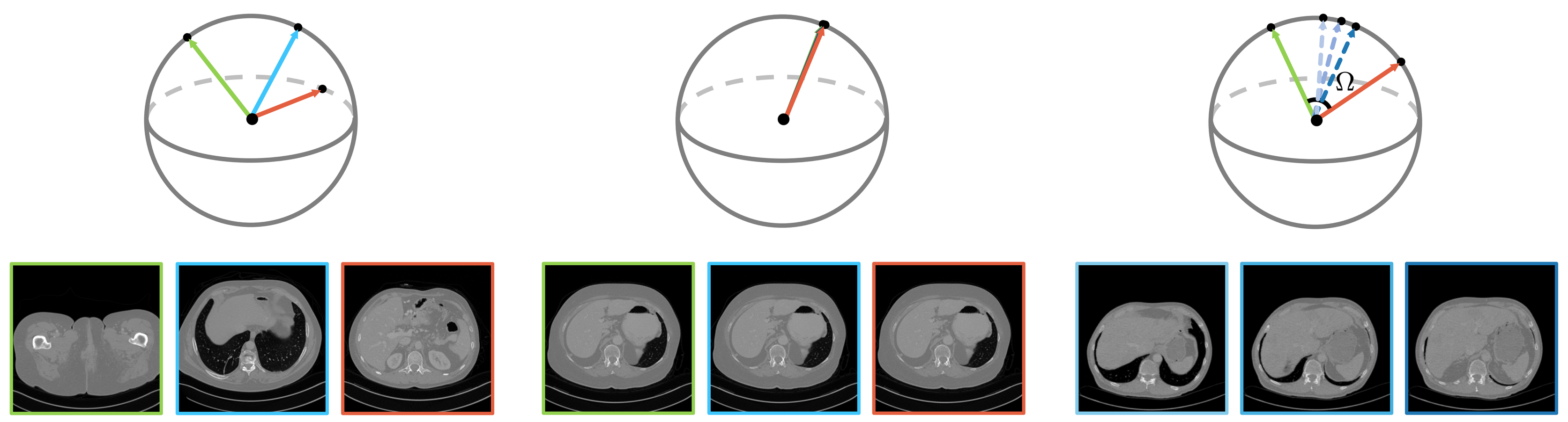
• How to build correlated noise volume?

An ideal correlated noise volume should have spatially varying correlations: (1) strong between adjacent slices for local consistency, and (2) gradually decaying with distance to allow global structural variation.

interpolation-based strategy: For a volume composed of S slices, sampling two anchor noise vectors (start and end points of a geodesic path on the hypersphere), $\mathbf{z}_1, \mathbf{z}_S$.

Then generate the correlated noise volume by Spherical Linear Interpolation *(Slerp)*

$$\epsilon_i^{\text{ISCS}} = \text{slerp}(\mathbf{z}_1, \mathbf{z}_S; \alpha_i) = \frac{\sin((1-\alpha_i)\Omega)}{\sin(\Omega)}\mathbf{z}_1 + \frac{\sin(\alpha_i\Omega)}{\sin(\Omega)}\mathbf{z}_S$$



(a) Independent noise (b) Identical noise (BCS) (c) Slerp noise (Ours)

• How to integrate with diffusion samplers?

Just replace the independently noise in any diffusion sampler's re-noising step with the generated correlated noise volume.

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{x}}_{0|t} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\cdot\epsilon_\theta(\mathbf{x}_t) + \sigma_t\cdot\epsilon^{\text{ISCS}}.$$

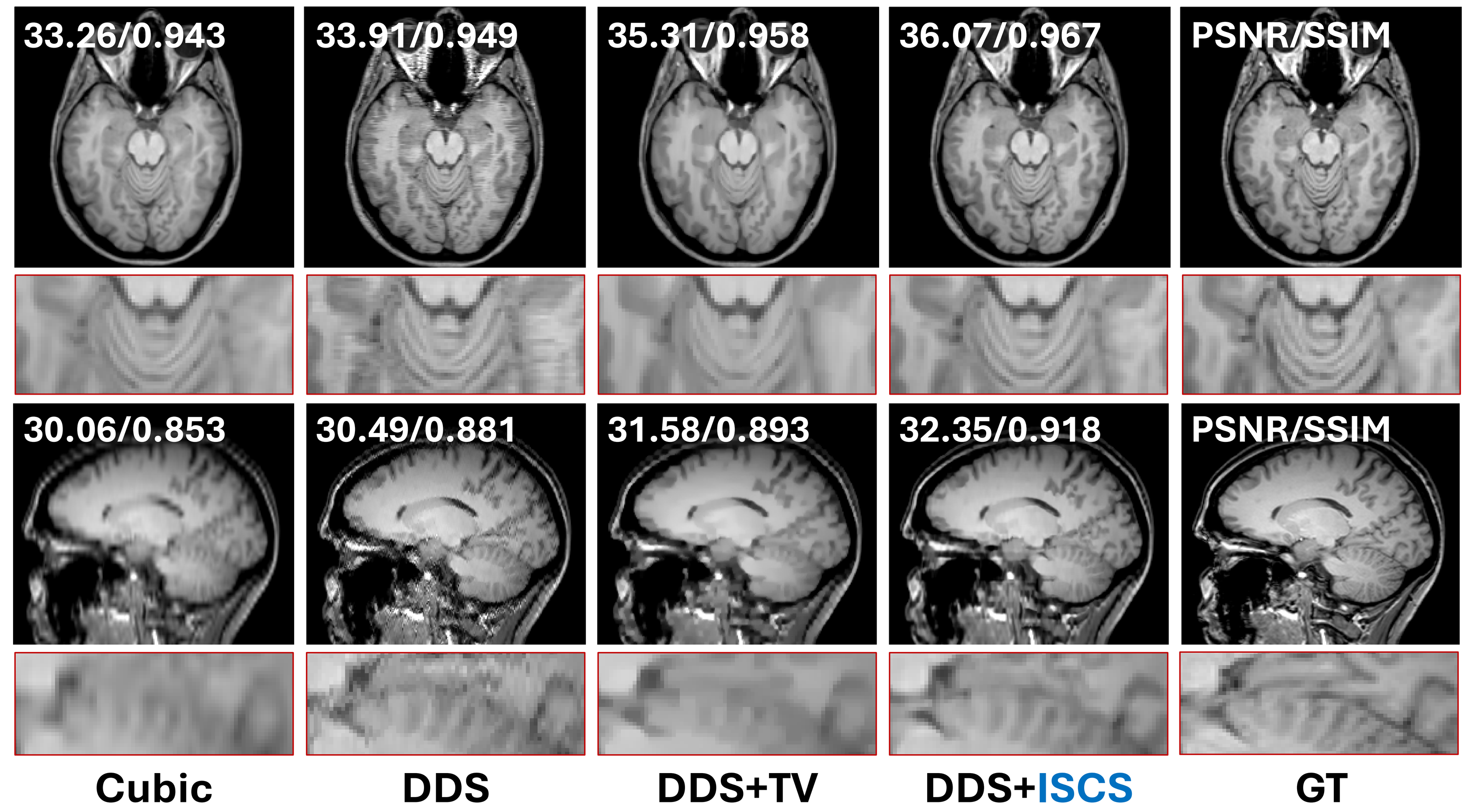
Algorithm 1 2D DIS for 3D medical imaging with ISCS

Input: $\mathbf{y}$, Pre-trained DM ϵ_{θ^*}, Timesteps T, Noise schedule $\{\alpha_t\}_{t=0}^T$.
$\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$;
for $t = T-1, \dots, 0$ **do**
 ▷ *1. Denoising Prediction*
 $\mathbf{x}_{0|t} = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_{\theta^*}(\mathbf{x}_t))$
 ▷ *2. Data Fidelity Update*
 $\hat{\mathbf{x}}_{0|t} = \arg\min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z} - \mathbf{x}_{0|t}\|_2^2$
 ▷ *3. Re-noising via ISCS*
 $\epsilon_i^{\text{ISCS}} = \texttt{slerp}(\mathbf{z}_1, \mathbf{z}_S; \alpha_i), \mathbf{z}_1, \mathbf{z}_S \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I})$
 $\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{x}}_{0|t} + \sqrt{1-\bar{\alpha}_{t-1}-\eta^2\tilde{\beta}_t^2}\epsilon_{\theta^*}^{(t)}(\mathbf{x}_t) + \eta\tilde{\beta}_t\epsilon^{\text{ISCS}}$
end for
return $\mathbf{x}_0$

Drop-in for Any DIS, No Optimization, No Extra Cost.

Results

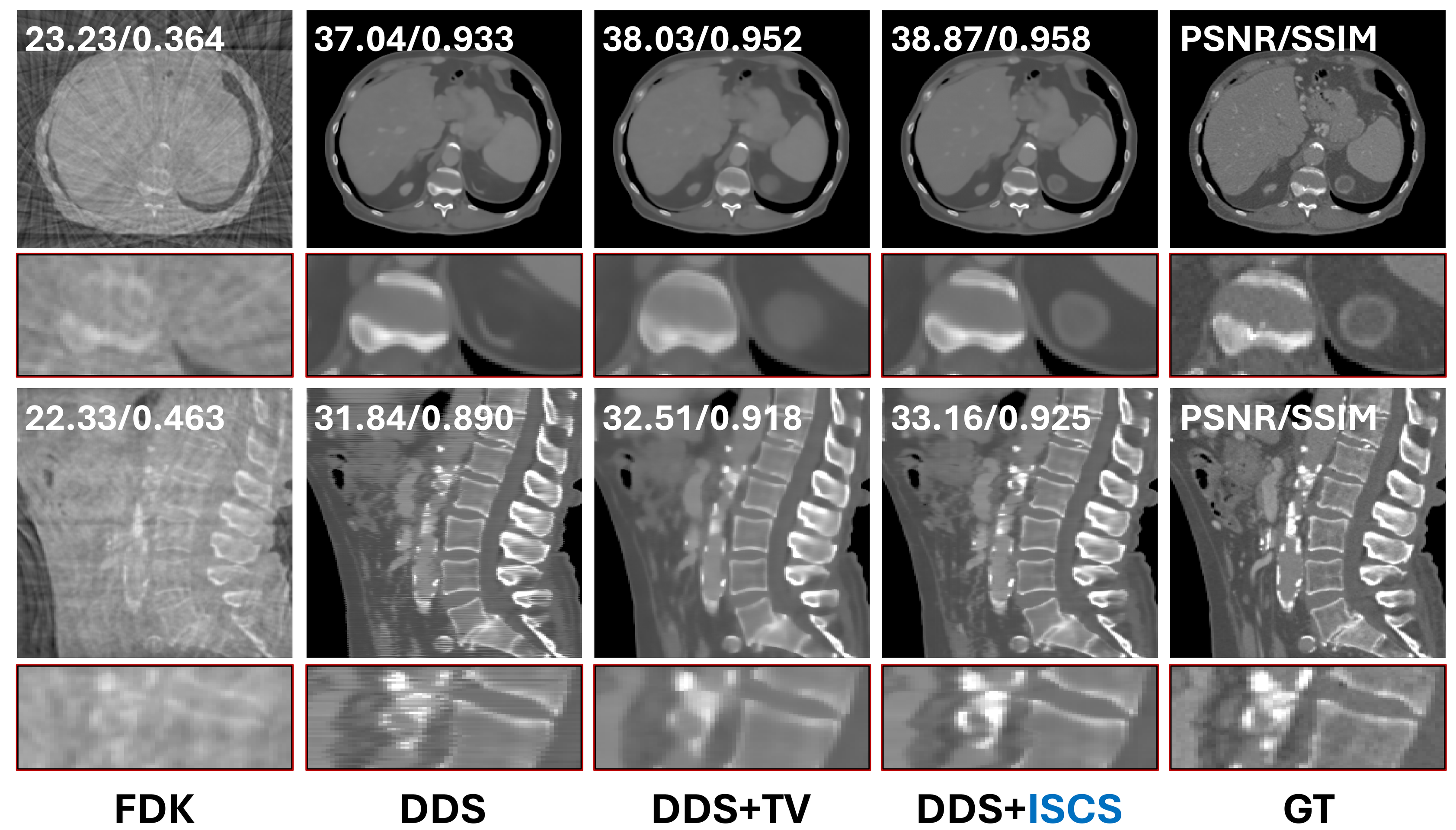
■ Experiments on *MRI SR 5×*



Total Variation (Post-Hoc)
- Enforces smoothness directly on the result
- Mitigates the **visible effects** of stochasticity

ISCS (source-level control)
- Controls stochasticity during sampling
- Targets the **root cause** of stochasticity

■ Experiments on *SVCT of 30 views*



■ Trajectory of the sampling process